DOCUMENT RESUME

ED 252 750 CG 017 918

AUTHOR Messana, Susan Melissa
TITLE Analysis of the Family Therapist Coding System during Ongoing Marital Therapy Sessions: Is It Reliable?
PUB DATE Mar 84
NOTE 15p.; Paper presented at the Annual Meeting of the Southeastern Psychological Association (30th, New Orleans, LA, March 28-31, 1984).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Counselors; *Interaction Process Analysis; *Interrater Reliability; *Marriage Counseling; *Verbal Communication
IDENTIFIERS *Family Therapist Coding System

ABSTRACT

Coding systems have become popular methods of cataloging the verbal and nonverbal interaction occurring during marital and family therapy. One such system, Pinsof's (1981) Family Therapist Coding System (FTCS), was the first designed explicitly to identify and differentiate specific verbal behaviors of family therapists independent of their theoretical orientation. To test the system's interrater and intrarater reliability, data were coded from typed manuscripts of six audio-taped marital therapy sessions. Coders were two undergraduate students trained for about seven hours each. The code consisted of three categories (for verb, phrase, and speech clause); the codes are ranked hierarchically such that only one code is assigned to each of the nine scales within the categories. The results indicated low observer agreement for overall session reliability and for category reliability. The low reliability did not appear to be due to observer drift or actual therapy sessions, but to the expertise and experience of the coders. The primary contributor to low reliability appeared to be the individual codes. Most of the codes with lower percentage of agreement values appeared to be less clearly defined and more difficult to apply to the data. The FTCS does not appear to be a reliable or practical assessment tool for determining the effectiveness of the therapist's statements during ongoing marital therapy sessions. (LLL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED252750

ANALYSIS OF THE FAMILY THERAPIST CODING SYSTEM DURING ONGOING MARITAL THERAPY SESSIONS: IS IT RELIABLE?

Susan Melissa Messana

Auburn University

Fall, 1984

CG 017918

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Susan M. Messana

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Introduction

Coding systems have become popular methods of cataloging the verbal and nonverbal interaction occurring during marital and family therapy. Although it appears that such techniques are necessary for the study of the process in marital and family interaction, few replications of research involving coding systems have been conducted and only a small percentage of these research studies have reported reliability statistics. Most research on marriage and family coding systems have reported overall interrater reliability in terms of percentage of agreement. Few of the studies gave specific, detailed descriptions of the sampled behaviors or the scoring unit, and failed to state which codes were reliable and which were not.

In order to assess the usefulness and accuracy of therapy coding systems, thorough studies must be conducted to determine the reliability of their application since conclusions cannot be drawn nor hypotheses tested until these coding systems are found to be both reliable and valid.

Purpose

The purpose of this study was to test the reliability of an application of Pinsof's (1981) Family Therapist Coding System (FTCS) in actual ongoing marital therapy sessions. In planning this study, the decision was made to analyze the therapist's statements during the therapy process. The effectiveness of specific therapist statements are critical to the therapy process and behavior change in the clients. The decision to use the FTCS was based on the belief that the FTCS is the most complex and reliable system developed thus far to describe a therapist's interaction in marital and family therapy.

The Family Therapist Coding System

The Family Therapist Coding System (FTCS) was the first coding system designed explicitly to identify and differentiate specific verbal behaviors of family therapists independent of their theoretical orientation. The FTCS consists of 9 nominal scales each of which contains numerous qualitatively distinct categories and sub-categories. In addition, the therapist's verbal behavior is coded within the context of the therapy interaction; that is, client statements can be used to clarify the therapist's statement. The FTCS is applied to written transcripts of therapy sessions and therefore, allows unitization prior to the coding process.

Statistics

While the most commonly used statistic for non-parametric data is percentage of agreement, this measure has several problems: (1) it does not take into account the chance occurrence of agreement, thereby resulting in a high reliability estimate; (2) percentage of agreement does not have metric properties and therefore, comparisons with other statistical measures are not possible; (3) percentage scores do not provide information about the sources of measurement error (e.g., errors of commission vs. errors of omission); (4) since percentage of agreement varies with the size of the time/event interval used, percentage of agreement scores are unrealistic when the rate of behavior is either very low or very high. (5) Finally, it is difficult to put percentage of agreement differences in perspective without knowledge of within subject variability. In the present study, in addition to adjusted percentage of agreement, Cohen's Kappa was applied to the data in order to enhance our understanding and knowledge of the coding scale's reliability.

Cohen's Kappa provides a superior statistic for reliability of non-parametric data since it accounts for the frequency with which coders use

each category in the scale and also, the extent to which a score differs from chance (Hollenbeck, 1978). Further, Kappa has a number of advantages for use with nominal scale data; one, it is easy to compute and two, it is valuable in training coders since it allows one to see each category and determine where the disagreements and agreements occur, thus, making differences in scoring easy to detect. Finally, Cohen's Kappa, unlike percentage of agreement, has metric properties which permit comparison of the results. One disadvantage of Cohen's Kappa is that the derived value decreases with increased amounts of data; thus, the important results may tend to be suppressed.

Questions

Is the Family Therapist Coding System reliable in terms of session reliability and category reliability?

1) Session Reliability
   a) What effects do the sequence of therapy sessions, the individual differences in coders, the experience and expertise of the coders, coder drift, and the order in which the therapy sessions were coded have on interrater reliability?
   b) What effects do the sequence of therapy sessions, the order in which the therapy sessions were coded, and coder drift have on intrarater reliability?
2) Category Reliability
   a) What effects do the three primary categories have on the reliability?
   b) What effects do the matches and non-matches of the individual codes have on the reliability?

Method

Two undergraduate students and the researcher coded data from typed manuscripts of audio-taped therapy sessions. Six sessions were coded for interrater reliability. These sessions were chosen to sample the range of the therapy sequence and were randomly assigned to the coding sequence. Prior to coding, each verb, phrase and speech clause was delineated, so that each of the coders would code each clause within the correct category. The codes are ranked heirarchically such that only one code is assigned to each of the 9 scales within the 3 categories. In order to obtain a measure of intrarater reliability, one of the coders recoded two-thirds of five previously coded sessions, in the same order as before. The two coders were trained for approximately seven hours each. Each coder was checked for accuracy against the researcher several times throughout the coding process.

Results

The data were analyzed by means of a computer program designed to calculate Cohen's Kappa and adjusted percentage of agreement for the overall session and individual category error rate (CRESCAT: Software for real-time analysis, 1981). In addition, the CRESCAT program designated the percentage of error for each category as well as the specification and location of the disagreement. Analyses were run for each coding pair for each of the therapy sessions.

The results indicate that in terms of the sequence of therapy sessions and the order in which the sessions were coded, there is little change in interobserver agreement and that the agreement obtained is not more than what would be expected by chance alone. In terms of all three coding categories, interobserver agreement was below .66 (Percentage of Agreement) and .24 (Cohen's Kappa).

The retest-reliability analyses indicate some increase in reliability over time, according to the order in which the sessions were coded. Also, the retest-reliability scores are higher than would be expected by chance alone, for the speech clause category and for the phrase category, after the first two sessions.

Looking at the individual codes, Table 1 shows the number of matches and non-matches plus the percentage of agreement for each code within the phrase category. The phrase category is divided into seven units; only one code from each scale is applied to each phrase clause. Only seven codes had above 50% agreement; the remainder of the scores were quite low.

Table 2 shows the individual code statistics for the verb category. Table 3 shows the individual code statistics for the speech clause category. One code I (Isolate) reached .82. The remaining 5 codes were quite low, below .20.

Discussion

The results indicate low observer agreement for both overall session reliability and category reliability. The low reliability does not appear to be due to observer drift or the actual therapy sessions. However, the expertise and experience of the coders did appear to affect the reliability results.

The primary contributer to the low reliability appears to be the individual codes; only nine codes within the three categories received agreement above 50%. The verb codes show the lowest reliability results, with only two codes showing percentage of agreement above 50%. The phrase category shows somewhat higher reliability results, in 5 of the 7 scales. This may be due, in part, to the fact that most of the scales contain fewer codes, in comparison to the verb category. The low reliability results shown in the

Intervention scale may be due to the effects of coder expertise. That is, only the researcher was knowledgeable of therapist intervention techniques, prior to the beginning of the study. In addition, the Intervention scale did not adequately represent or depict the range and specificity of the interventions contained in the sessions coded for this study.

The speech clause category, alone, shows a slightly higher rate of agreement that expected by chance as well as a slight tendency to increase in reliability across sessions. This may be due to the fact that the speech clause category contains much fewer codes than the other 2 categories. Also, the speech clause codes are applied to much less data.

The codes that show higher percentage of agreement values in all three categories appear to be more easily differentiated from the other codes. Most of the codes with lower percentage of agreement values appeared to be less clearly defined and more difficult to apply to the data.

Implications

In conclusion, the FTCS does not appear to be a reliable or practical assessment tool for determining the effectiveness of the therapist's statements during ongoing marital therapy sessions. The Intervention codes, which would appear to be the most useful scale in this regard, did not have any percentage of agreement scores above 38%. The time involved in training the coders and in preparing the data for coding limit the practicality of the FTCS. Future studies should have each coder apply a different coding scale, which has already been shown to be reliable, to dummy tapes, prior to beginning the actual study, to insure that the low reliability is not due to the individual coders.

TABLE 1

MATCHES, NON-MATCHES, AND PERCENTAGE OF AGREEMENT
FOR EACH 'PHRASE' CODE CATEGORY **

	CODE	#	MATCHES	NON-MATCHES	PERCENTAGE OF AGREEMENT
1.	DD	1	00	05	.00
	SD	2	58	56	.30
	R	3	02	02	.25
	DN	4	28	78	.38
	TR	5	00	04	.00
	EM	6	52	173	.19
	S	7	162	204	.36
	BR	8	00	10	.00
	EN	9	00	08	.00
	PR	10	00	16	.00
	C	11	140	212	.31
	PM	12	00	23	.00
	ST	13	36	150	.14
2.	N	14	158	274	.43
	CR	15	410	316	.55
	F	16	18	34	.34
	P	17	60	102	.20
	AT	18	00	15	.00
3.	CT	19	--	--	--
	FM	20	--	--	--
	PC	21	--	--	--
	CP	22	42	88	.24
	WM	23	462	163	.70
	HF	24	468	188	.70
	C+	25	--	--	--
	C1	26	--	--	--
	C2	27	--	--	--
	NS	28	--	--	--
4.	G	29	--	--	--
	TC	30	76	126	.29
	DC	31	182	224	.39
	MC	32	252	228	.49
	T	33	262	370	.42

** M01 Phrase Clauses

1. Intervention scale
2. Temporal Orientation scale
3. To Whom scale
4. Interpersonal Structure scale

TABLE 1 (CONTINUED)

MATCHES, NON-MATCHES, AND PERCENTAGE OF AGREEMENT
FOR EACH 'PHRASE' CODE CATEGORY **

	CODE	#	MATCHES	NON-MATCHES	PERCENTAGE OF AGREEMENT
5.	INC	34	00	38	.00
	TY	35	--	--	--
	CG	36	00	09	.00
	TFO	37	00	03	.00
	CFO	38	00	09	.00
	MFO	39	04	01	.80
	PFO	40	00	08	.00
	SFO	41	00	15	.00
	EFO	42	00	01	.00
	NFO	43	--	--	--
	TNF	44	14	114	.07
	CNF	45	00	07	.00
	MNF	46	440	216	.63
	PNF	47	00	06	.00
	SNF	48	00	08	.00
	ENF	49	02	02	.33
	NNF	50	04	42	.04
	OTF	51	04	29	.08
	OT	52	02	29	.07
	OF	53	02	20	.03
	O	54	00	10	.00
6.	IT	55	00	06	.00
	DI	56	02	37	.01
	I	57	00	03	.00
	D	58	947	69	.94
7.	CD	59	04	23	.12
	QO	60	134	131	.47
	QC	61	126	149	.46
	L	62	670	166	.77
8.	AA	63	00	25	.00
	INCL	64	00	06	.00

** MO1 Phrase Clauses

5. System Membership scale
6. Route scale
7. Grammatical Form scale
8. Event Relationship scale

TABLE 2

MATCHES, NON-MATCHES, AND PERCENTAGE OF AGREEMENT
FOR EACH 'VERB' CODE CATEGORY **

CODE	#	MATCHES	NON-MATCHES	PERCENTAGE OF AGREEMENT
CON	1	68	125	.32
PE	2	16	60	.09
NE	3	56	96	.36
NSE	4	74	88	.36
PB	5	160	328	.23
NB	6	46	162	.14
SB	7	00	06	.00
NVB	8	00	51	.00
VB	9	426	300	.51
NSB	10	500	723	.36
PC	11	00	62	.00
NVC	12	00	11	.00
NLC	13	1058	751	.54
SP	14	00	26	.00
EX	15	02	16	.00
F	16	00	29	.00
INCL	17	00	02	.00

** V01 Verb Clauses

TABLE 3

MATCHES, NON-MATCHES, AND PERCENTAGE OF AGREEMENT
FOR EACH 'SPEECH CLAUSE' CODE CATEGORY

CODE	#	MATCHES	NON-MATCHES	PERCENTAGE OF AGREEMENT
CY	1	00	10	.00
FN	2	16	61	.08
IN	3	00	04	.00
T	4	--	--	--
M	5	66	157	.18
IS	6	820	190	.82
INCL	7	06	38	.15

REFERENCES

Abramowitz, S. I., & Jackson, C. (1974). Comparative effectiveness of there-and-then versus here-and-now therapist interpretations in group psychotherapy. Journal of Counseling Psychology, 21, 288-293.

Allred, G. H. & Kerse', F. L. (1977). The AIAC, a design for systematically analyzing marriage and family counseling: A progress report. Journal of Marriage and Family Counseling, 19, 17-24.

Arkowitz, H., Lamke, L. K., & Filsinger, E. E. (1981). Issues of behavioral assessment: Final reflections. In E. E. Filsinger & R. A. Lewis (Eds.), Assessing marriage: New behavioral approaches. Beverly Hills: Sage Publications.

Bakeman, R. (1978). Untangling streams of behavior: Sequential analyses of observation data. In G. P. Sackett (Ed.), Observing behavior (vol. 2): Data collection and analysis methods. Baltimore: University Park Press.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 561-569.

Becker, J. (1974). Depression: Theory and research. New York: John Wiley and Sons.

Bellack, Alan S. & Hersen, Michael. (1977). Self-report inventories in behavioral assessment. In J. D. Cone & R. P. Hawkins (Eds.), Behavioral assessment: New directions in clinical psychology. New York: Bruner/Mazel.

Bradley, James V. (1972). Nonparametric statistics. In R. E. Kirk (Ed.), Statistical issues: A reader for the behavioral sciences. Monterey, Calif.: Brooks/Cole.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley and sons.

Dowling, E. (1979). Co:Therapy: A clinical researcher's view. In S. Walrond-Skinner (Ed.), Family and marital psychotherapy. London: Routledge and Kegan Paul.

Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of Kappa and weighted Kappa. Psychological Bulletin, 72, 323-327.

Gottman, J. (1979). Marital interaction: Experimental investigations. New York: Academic Press.

Gurman, A. S. & Kniskern, D. P. (1978). Research on marital and family therapy: Progress, perspective, and prospect. In S. L. Garfield & A. E. Bergin (Eds.), Handbook of psychotherapy and behavior change. New York: John Wiley and Sons.

Hartmann, D. P., & Gardner, W. (1981). Considerations in assessing the reliability of observations. In E. E. Filsinger & R. A. Lewis (Eds.), Assessing marriage: New behavioral approaches. Beverly Hills: Sage Publications.

Hawkins, R. P. & Fabry, B. D. (1979). Applied behavior analysis and inter-observer reliability: A commentary on two articles by Birkimer and Brown. Journal of Applied Behavior Analysis, 12, 545-552.

Hollenbeck, A. R. (1978). Problems of reliability in observational research. In G. P. Sackett (Ed.), Observing behavior (vol. 2): Data collection and analysis methods. Baltimore: University Park Press.

Johnson, S. M., & Bolstad, O. D. (1973). Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, & E. J. Mash (Eds.), Behavior change: Methodology, concepts and practice. Champaign, Illinois: Research Press.

Jones, R. R., Reid, J. B., & Patterson, G. R. (1975). Naturalistic observation in clinical assessment. In P. McReynolds (Ed.). Advances in psychological assessment (vol. 3). San Francisco: Jossey-Bass.

Kaye, K. (1980). Estimating false alarms and missed events from interobserver agreement: A rationale. Psychological Bulletin, 88, 458-468.

Kazdin, A. E. (1977). Artifact, bias and complexity of assessment: The ABC's of reliability. Journal of Applied Behavior Analysis, 10, 141-150.

Kratochwill, T. R. (1979). Just because it's reliable doesn't mean it's believable. Journal of Applied Behavior Analysis, 12, 553-557.

Markman, H. J., Notarius, C. I., Stephen, T., & Smith, R. J. (1981). Behavioral observation systems for couples: The current status. In E. E. Filsinger & R. A. Lewis (Eds.), Assessing marriage: New behavioral approaches. Beverly Hills: Sage Publications .

Mintz, J. & Luborsky, L. (1971). Segments versus whole sessions: Which is the better unit for psychotherapy process research? Journal of Abnormal Psychology, 78, 180-191.

Mitchell, S. K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. Psychological Bulletin, 86, 376-390.

Notarius, C. I., Krokoff, L. J. & Markman, H. J. (1981). Analysis of Observational Data. In E. E. Filsinger & R. A. Lewis (Eds.), Assessing marriage: New behavioral approaches. Beverly Hills: Sage Publications.

Olson, David H. (1981). Family typologies: Bridging family research and family therapy. In E. E. Filsinger & R. A. Lewis (Eds.), Assessing marriage: New behavioral approaches. Beverly Hills: Sage Publications.

Orlinsky, D. E. & Howard, K. I. (1978). The relation of process to outcome in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), Handbook of psychotherapy and behavior change. New York: John Wiley and Sons.

Pinsof, W. M. (1981). Family therapy process research. In A. S. Gurman & D. P. Kniskern (Eds.), Handbook of family therapy. New York: Brunner/ Mazel.

Pinsof, W. M. (1979). The family therapist behavior scale (FTBS): Development and evaluation of a coding system. Family Process, 18, 451-461.

Pinsof, W. M. (1981). The family therapist coding system (FTCS) coding manual, Unpublished manuscript, Center for Family Studies, Department of Psychiatry and Behavioral Sciences, Northwestern University Medical School.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in general population. Applied Psychological Measurement, 1, 385-401.

Rehm, L. P. (1981). Assessment of depression. In M. Hersen & A. S. Bellack (Eds.), Behavioral assessment: A practical handbook. New York: Pergamon Press.

Riskin, J. & Faunce, E. E. (1972). An evaluative review of family interaction research. Family Process, 11, 365-455.

Sackett, G. P., Ruppenthal, G. C., & Gluck, J. (1978). Introduction: An overview of methodological and statistical problems in observational research. In G. P. Sackett (Ed.), Observing behavior (vol. 2:) Data collection and analysis methods. Baltimore: University Park Press.

Taplin, P. S., & Reid, J. B. (1973). Effects of instructional set and experimenter influence on observer reliability. Child Development, 3, 261-271.

Waxler, N. E. & Mishler, E. G. (1966). Scoring and reliability problems in interaction process analysis: A methodological note. Sociometry, 29, 28-40.

Weissman, M. M., Sholomskas, D., Pottenger, M., Prusoff, B. A., & Locke, B. Z. (1977). Assessing depressive symptoms in five psychiatric populations: A Validation Study. American Journal of Epidemiology, 106, 203-214.

Wildman, B. G. & Erickson, M. T. (1977). Methodological problems in behavioral observation. In J. D. Cone & R. P. Hawkins (Eds.), Behavioral assessment: New directions in clinical psychology. New York: Bruner/Mazel.